STOCHASTIC REDUCTION METHOD USING TIME-SCALE SEPARATION IN CHEMICAL KINETIC SYSTEMS AND REGULATORY NETWORKS

CHETAN D. PAHLAJANI, PAUL J. ATZBERGER AND MUSTAFA KHAMMASH

1. Introduction

Recent years have seen an explosion of interest in the mathematical analysis of cellular and sub-cellular biological systems. (Sentence or two on why, larger field of systems biology, experimental advances, etc.)

At the heart of biological systems analysis is the analysis of *biochemical reaction networks*. Understanding chemical kinetics at the cellular level poses unique challenges not traditionally encountered at the macroscopic (test-tube) scale. Perhaps the most significant of these is the presence of small numbers of molecules of reactants and products (this is especially true in gene transcription networks where there may be tens or hundreds of molecules of some species). The result is an inherent, observable stochasticity that cannot be accounted for by traditional deterministic models. A second challenge is the presence of chemical reactions occurring on vastly different time scales.

Much of the interest, then, has been devoted to developing and analyzing stochastic models of cellular processes. Assuming well-mixedness of the cell contents - a reasonable assumption under a wide range of conditions - one should study chemical reactions within the framework of *continuous-time Markov chains*. In such descriptions, one keeps track of populations of reactants and products, i.e. copy numbers . These Markov chains can be mathematically characterized using random time changes (see [EK86]) or in terms of the chemical master equation which describes time evolution of probabilities (see [Gar04], [vK]). Much recent work has focused on direct simulation, both exact and approximate, of the trajectories of these Markov chains. This involves using the stochastic simulation algorithm (SSA), also known as Gillespie's algorithm, and its variants (see [SKPG] for a survey and references). A different approach to approximating solutions to the chemical master equation (CME) is the finite-state projection approach described in [MK06]. The importance of these techniques stems from the fact that exact simulation can be computationally expensive, especially in the presence of multiple time scales.

Another important approximation technique is the use of the *linear noise approximation* (LNA) or *van Kampen approximation* (see [vK], [EE03]). Here, instead of working with the chemical master equation, the Markov chain is approximated by the macroscopic ordinary differential equation (ODE) of classical chemical kinetics, together with a Gaussian fluctuation process described by a stochastic differential equation (SDE). The LNA is thus a mesoscopic model - intermediate between the microscopic Markov chain and the macroscopic ODE. The main advantages of the LNA formulation are that it admits stochasticity in the model while working at a coarse enough scale, and that the powerful highly-developed machinery of stochastic calculus can be brought to bear on the problem at hand.

We consider here the linear noise approximation for a chemically reacting system with two distinct time scales. The resulting mathematical model comprises a multiscale ODE driving a multiscale SDE. This opens the door to applying reduction techniques for singularly perturbed ODE and SDE

Date: October 15, 2009.

(see [Pap76, Kur73, PS08, MTV]). Indeed, our main result is an asymptotic model reduction of precisely this type.

The paper is organized as follows. In Section 2, we describe an autoregulatory single gene network that serves as a prototype for the general chemical kinetics system described in Section 3. In Section 3, we describe our general system and quickly state our model reduction result. Section 4 is devoted to the calculations that justify our result in Section 3. Finally, in Section 5, we return to the single gene network to illustrate our results and also provide numerical validation in this special case.

2. Background

To motivate the general chemical kinetics systems to be studied in the sequel, we start with a relatively simple, yet frequently encountered example of a gene network. This is the case of an autoregulatory gene network consisting of a single gene whose protein inhibits its own production.

For our purposes, it will suffice to think of a gene as a specific stretch of DNA that encodes for a specific protein. By gene expression, we mean the sequence of chemical reactions through which a gene (call it Gene_1) makes the corresponding protein (call it $\mathsf{Protein}_1$). While the mechanisms underlying the expression of a gene are undoubtedly rather complex, a frequently used model is the following: RNA polymerase (RNAp) in the cell binds to a regulatory region preceding the gene called the promoter and produces mRNA transcripts (in a process called transcription) corresponding to the gene's coding sequence. These mRNA transcripts are then translated into the corresponding protein (in a process called translation). The sequence of chemical reactions can be summarized as follows:

$$\begin{aligned} \mathsf{Gene}_1 &\to \mathsf{Gene}_1 + \mathsf{mRNA}_1 \\ \mathsf{mRNA}_1 &\to \emptyset \\ \mathsf{mRNA}_1 &\to \mathsf{mRNA}_1 + \mathsf{Protein}_1 \\ \mathsf{Protein}_1 &\to \emptyset; \end{aligned} \tag{1}$$

here $\emptyset$ indicates decay of the molecule.

The rate at which RNAp initiates transcription for a specific gene is regulated by special proteins called transcription factors. Indeed, the latter bind specific sites on the promoters and can increase or decrease the rate at which mRNA transcripts are produced and can thus serve as, respectively, activators or inhibitors. A cell typically contains several thousand types of proteins. The result is a network structure with genes as nodes and a directed edge taking Gene_1 to Gene_2 signifying that the protein product of Gene_1 is a transcription factor regulating the mRNA (and thus protein) production of Gene_2 (see [A07]). In the case of an autoregulatory gene network, a protein is a transcription factor for its own production.

Figure 1 depicts the autoregulatory gene network we have in mind. The quantities next to the arrows denote the rates at which various reactions occur. The rate of transcription is described by a Hill function with Hill coefficient N, maximum rate of transcription k_r^{max} and repression coefficient K_D. The rate of decay of the mRNA and the rate of production of the protein are, respectively, γ_r times the number of mRNA molecules, and k_p times the number of mRNA molecules. The rate of decay of protein is γ_p times the number of protein molecules.

Replacing k_r^{max} by k_r^{max}/ε and γ_r by γ_r/ε where $0 < \varepsilon \ll 1$ is a small parameter, we get the prototype of the multiscale chemical kinetics system studied in this paper. This rescaling corresponds (in the single gene case) to the assumption that the reactions involving mRNA production and degradation occur on a much faster time scale than the reactions involving protein production and degradation. (Justify this.)

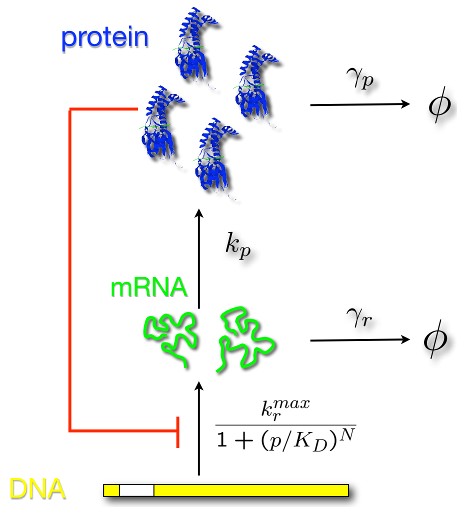

FIGURE 1. An autoregulatory gene network with negative feedback.

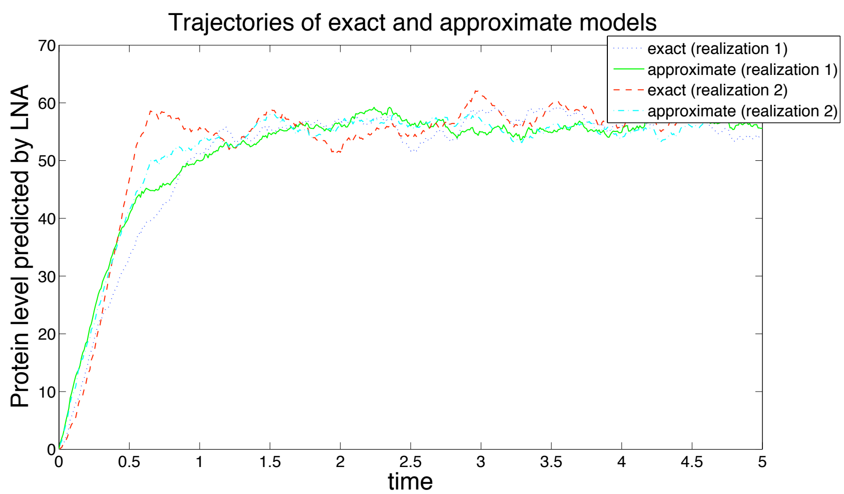

FIGURE 2. Sample paths for the autoregulatory gene network with the LNA. The parameters in the problem (see Section 5) were $k_r^{max} = 10$, $\gamma_r = 1$, $k_p = 10$, $\gamma_p = 1$, $N = 4$, $K_D = 60$, $n = 25$.

Some sample trajectories for this system (within the framework of the LNA) together with the trajectories for the reduced system are shown in figure 2.

Now that we have a concrete example in mind, we proceed in the next section to describe the problem and statement of results for a general chemical kinetics system. In section 5, we come full circle and illustrate our results and calculations of sections 3 and 4 in the context of the single gene autoregulatory network of Figure 1.

3. Problem Statement and Main Result

3.1. **Underlying Markov chain and the linear noise approximation.** We start by describing a continuous-time Markov chain $\mathbf{X}^{n,\varepsilon}(t) \in (\mathbb{Z}^+)^N$ that gives the population vector for a chemically reacting system with N species. Here, n is the reaction volume and $0 < \varepsilon \ll 1$ is a small parameter that captures time scale separation. We outline the approximation (as $n \to \infty$, $\varepsilon \in (0,1)$ fixed) of the Markov chain by the ordinary differential equation (ODE) (10) and write the stochastic differential equation (SDE) (12) that approximates the rescaled fluctuations of the Markov chain from the ODE. These $n \to \infty$ asymptotic calculations are classical (see Chapter 11 in [EK86] and references therein). The result is now an ε-dependent system - a singularly perturbed ODE driving a singularly perturbed SDE. The ensuing $\varepsilon \to 0$ asymptotics are the main content of this work.

As stated above, we have N chemical species $S_1, S_2, \dots, S_N$, reacting through M reaction channels $R_1, R_2, \dots, R_M$. We will denote the population vector of the N species by $\mathbf{X}^{n,\varepsilon}(t) \stackrel{\text{def}}{=} (X_1^{n,\varepsilon}(t), X_2^{n,\varepsilon}(t), \dots, X_N^{n,\varepsilon}(t)) \in (\mathbb{Z}^+)^N$. Let $\nu_1, \nu_2, \dots, \nu_M$ be vectors in $\mathbb{Z}^N$ which are the stoichiometric vectors for the M reactions, i.e. a single occurrence of reaction R_k at time t changes the population vector from $\mathbf{X}^{n,\varepsilon}(t-)$ to $\mathbf{X}^{n,\varepsilon}(t) = \mathbf{X}^{n,\varepsilon}(t-) + \nu_k$. We write $\nu_k = (\nu_k^1, \nu_k^2, \dots, \nu_k^N)$.

We will partition the set of M reactions into slow reactions and fast reactions; essentially, the intensities (propensities) of the fast reactions will have a prefactor of $1/\varepsilon$ as will be elaborated below. Let M_{s}, M_{f} be positive integers such that $M_{\mathsf{s}} + M_{\mathsf{f}} = M$. We can assume, without loss of generality, that the reactions R_1 through $R_{M_{\mathsf{s}}}$ are slow, while reactions $R_{M_{\mathsf{s}}+1}$ through R_M are fast. Let

$$\Lambda_{\mathsf{s}} \stackrel{\text{def}}{=} \{1, 2, \dots, M_{\mathsf{s}}\} \qquad \text{and} \qquad \Lambda_{\mathsf{f}} \stackrel{\text{def}}{=} \{M_{\mathsf{s}}+1, M_{\mathsf{s}}+2, \dots, M\}. \tag{2}$$

Let's now specify the intensities for the various reactions. For $k = 1, 2, \dots, M$, fix smooth functions $\beta_k : \mathbb{R}^N \to [0, \infty)$ such that $\beta_k(\mathbf{x}) \geq 0$ whenever $\mathbf{x} = (x_1, x_2, \dots, x_N) \in (\mathbb{R}^+)^N$. The intensities for the reactions are now given by

$$\beta_k^\varepsilon(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} \beta_k(\mathbf{x}) & \text{if } k \in \Lambda_{\mathsf{s}} \\ \frac{1}{\varepsilon}\beta_k(\mathbf{x}) & \text{if } k \in \Lambda_{\mathsf{f}} \end{cases} \tag{3}$$

for $\mathbf{x} \in (\mathbb{Z}^+)^N$. The evolution of the Markov chain $\mathbf{X}^{n,\varepsilon}(t)$ is described by the requirement that for $t \geq 0$, $1 \leq k \leq M$, we have, as $h \searrow 0$,

$$\mathbb{P}\left(\mathbf{X}^{n,\varepsilon}(t+h) = \mathbf{X}^{n,\varepsilon}(t) + \nu_k | \mathbf{X}^{n,\varepsilon}(s) : 0 \leq s \leq t\right) = n\beta_k^\varepsilon\left(\frac{\mathbf{X}^{n,\varepsilon}(t)}{n}\right) h + o(h). \tag{4}$$

This can be formalized by either writing the chemical master equation (CME) for the evolution of probabilities or by working with random time changes (Section 6.4 in [EK86]). We opt for the latter and require that $\mathbf{X}^{n,\varepsilon}(t)$ solve the random time change equation

$$\mathbf{X}^{n,\varepsilon}(t) = \mathbf{X}^{n,\varepsilon}(0) + \sum_{k=1}^{M} \nu_k Y_k \left(n \int_0^t \beta_k^\varepsilon\left(\frac{\mathbf{X}^{n,\varepsilon}(s)}{n}\right) ds \right) \tag{5}$$

where the Y_k's are independent unit rate Poisson processes. $\mathbf{X}^{n,\varepsilon}(t)$ is thus a *density dependent population process* (see [EK86]). Now,

$$\frac{\mathbf{X}^{n,\varepsilon}(t)}{n} = \frac{\mathbf{X}^{n,\varepsilon}(0)}{n} + \sum_{k=1}^{M} \nu_k \frac{1}{n} Y_k \left(n \int_0^t \beta_k^\varepsilon\left(\frac{\mathbf{X}^{n,\varepsilon}(s)}{n}\right) ds \right). \tag{6}$$

Let

$$F^\varepsilon(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{k=1}^{M} \nu_k \beta_k^\varepsilon(\mathbf{x}). \tag{7}$$

For $\varepsilon \in (0,1)$ fixed, if

$$\lim_{n\to\infty} \frac{\mathbf{X}^{n,\varepsilon}(0)}{n} = \mathbf{x}^{\mathsf{init}}, \tag{8}$$

then a form of the Law of Large Numbers (LLN) (Theorem 11.2.1 [EK86]) states that for $T < \infty$ arbitrary,

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left| \frac{\mathbf{X}^{n,\varepsilon}(t)}{n} - \mathbf{x}^\varepsilon(t) \right| = 0, \qquad \mathbb{P}\text{-a.s.} \tag{9}$$

where $\mathbf{x}^\varepsilon(t)$ solves the ODE

$$\begin{aligned} \dot{\mathbf{x}}^\varepsilon(t) &= F^\varepsilon(\mathbf{x}^\varepsilon(t)) \\ \mathbf{x}^\varepsilon(0) &= \mathbf{x}^{\mathsf{init}}. \end{aligned} \tag{10}$$

Moreover, for $\varepsilon \in (0,1)$ fixed, the rescaled fluctuations about the mean given by $\sqrt{n}(\mathbf{X}^{n,\varepsilon}(t)/n - \mathbf{x}^\varepsilon(t))$ satisfy a *central limit theorem* as $n \to \infty$ (see Theorem 11.2.3 in [EK86]). Indeed, if $\lim_{n\to\infty} \sqrt{n}(\mathbf{X}^{n,\varepsilon}(0)/n - \mathbf{x}^\varepsilon(0)) = \mathbf{v}^{\mathsf{init}}$ (constant), then $\sqrt{n}(\mathbf{X}^{n,\varepsilon}(t)/n - \mathbf{x}^\varepsilon(t)) \Rightarrow \mathbf{V}^\varepsilon(t)$ where $\Rightarrow$ denotes convergence in distribution and $\mathbf{V}^\varepsilon(t)$ is the time-inhomogeneous Gaussian process given by

$$\mathbf{V}^\varepsilon(t) = \mathbf{v}^{\mathsf{init}} + \int_0^t DF^\varepsilon(\mathbf{x}^\varepsilon(s))\mathbf{V}^\varepsilon(s)ds + \sum_{k=1}^{M} \nu_k \int_0^t \sqrt{\beta_k^\varepsilon(\mathbf{x}^\varepsilon(s))}dW_k(s), \tag{11}$$

with $W \stackrel{\text{def}}{=} (W_1, W_2, \dots, W_M)$ a standard M-dimensional Brownian motion on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In differential form, the equation (11) can be written

$$\begin{aligned} d\mathbf{V}^\varepsilon(t) &= DF^\varepsilon(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon(t)dt + \sum_{k=1}^{M} \nu_k \sqrt{\beta_k^\varepsilon(\mathbf{x}^\varepsilon(t))}dW_k(t) \\ \mathbf{V}^\varepsilon(0) &= \mathbf{v}^{\mathsf{init}}. \end{aligned} \tag{12}$$

The assertion that

$$\frac{\mathbf{X}^{n,\varepsilon}(t)}{n} \approx \mathbf{x}^\varepsilon(t) + \frac{1}{\sqrt{n}}\mathbf{V}^\varepsilon(t) \qquad \text{for large } n \tag{13}$$

constitutes the *linear noise approximation*. The pair of equations (10) and (12) will be the starting point of our asymptotic analysis. The significance of $\mathbf{x}^\varepsilon(t)$ and $\mathbf{V}^\varepsilon(t)$ from the standpoint of the original Markov chain $\mathbf{X}^{n,\varepsilon}(t)$ is encapsulated in (13).

3.2. **Main Result.** We are now in a position to clearly state our singular perturbations problem and describe our main result. We start with

Assumption 3.1. Assume that the N chemical species can be partitioned into slow and fast species as follows : there exists a positive integer $\ell < N$ such that species $S_1, S_2, \dots, S_\ell$ - the "slow "species - experience population changes *only* through the (slow) reactions R_1 through R_{M_s}, while species $S_{\ell+1}$ through S_N - the "fast "species - experience population changes *only* through the (fast) reactions $R_{M_\mathsf{s}+1}$ through R_M [1]. The β_k's, however, may depend on the populations of *both* the slow and fast species.

The equations (10), (12) form a singularly perturbed ODE driving a singularly perturbed SDE. To see this, we introduce a bit of notation. Let $\mathsf{P}_\mathsf{s} : \mathbb{R}^N \to \mathbb{R}^\ell$ and $\mathsf{P}_\mathsf{f} : \mathbb{R}^N \to \mathbb{R}^{N-\ell}$ be the projection operators given by

$$\mathsf{P}_\mathsf{s}(x_1, x_2, \dots, x_N) \stackrel{\text{def}}{=} (x_1, \dots, x_\ell) \qquad \text{and} \qquad \mathsf{P}_\mathsf{f}(x_1, x_2, \dots, x_N) \stackrel{\text{def}}{=} (x_{\ell+1}, \dots, x_N) \tag{14}$$

for $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$. We can now write a vector $\mathbf{x} \in \mathbb{R}^N$ as $\mathbf{x} = (\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f})$ where $\mathbf{x}_\mathsf{s} = \mathsf{P}_\mathsf{s}\mathbf{x}$ and $\mathbf{x}_\mathsf{f} = \mathsf{P}_\mathsf{f}\mathbf{x}$. Assumption 3.1 can be expressed as follows: for $\mathbf{x} = (\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f}) \in (\mathbb{Z}^+)^N$,

$$\mathbf{x} + \nu_k = \begin{cases} (\mathbf{x}_\mathsf{s} + \mathsf{P}_\mathsf{s}\nu_k, \mathbf{x}_\mathsf{f}) & \text{if } k \in \Lambda_\mathsf{s} \\ (\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f} + \mathsf{P}_\mathsf{f}\nu_k) & \text{if } k \in \Lambda_\mathsf{f}. \end{cases} \tag{15}$$

Next, we note that the Jacobian matrix DF^ε admits a nice block decomposition. For $g : \mathbb{R}^N \to \mathbb{R}^p$ smooth, let $D_\mathsf{s}g$ denote the Jacobian matrix of the function g with respect to the variables x_1 through x_ℓ, and let $D_\mathsf{f}g$ denote the Jacobian matrix of g with respect to the variables $x_{\ell+1}$ through x_N. Writing $F^\varepsilon = (F_\mathsf{s}, \frac{1}{\varepsilon}F_\mathsf{f})$ (in other words, $F_\mathsf{s} \stackrel{\text{def}}{=} \mathsf{P}_\mathsf{s}F^\varepsilon$ and $F_\mathsf{f} \stackrel{\text{def}}{=} \mathsf{P}_\mathsf{f}(\varepsilon F^\varepsilon)$), $\mathbf{x}^\varepsilon(t) = (\mathbf{x}^\varepsilon_\mathsf{s}(t), \mathbf{x}^\varepsilon_\mathsf{f}(t))$, $\mathbf{V}^\varepsilon(t) = (\mathbf{V}^\varepsilon_\mathsf{s}(t), \mathbf{V}^\varepsilon_\mathsf{f}(t))$, and the derivative matrix DF^ε in block form

$$DF^\varepsilon = \begin{bmatrix} D_\mathsf{s}F_\mathsf{s} & D_\mathsf{f}F_\mathsf{s} \\ \frac{1}{\varepsilon}D_\mathsf{s}F_\mathsf{f} & \frac{1}{\varepsilon}D_\mathsf{f}F_\mathsf{f} \end{bmatrix}, \tag{16}$$

the equations (10), (12) become, respectively,

$$\begin{aligned} \dot{\mathbf{x}}^\varepsilon_\mathsf{s}(t) &= F_\mathsf{s}(\mathbf{x}^\varepsilon_\mathsf{s}(t), \mathbf{x}^\varepsilon_\mathsf{f}(t)) \\ \dot{\mathbf{x}}^\varepsilon_\mathsf{f}(t) &= \frac{1}{\varepsilon}F_\mathsf{f}(\mathbf{x}^\varepsilon_\mathsf{s}(t), \mathbf{x}^\varepsilon_\mathsf{f}(t)) \\ (\mathbf{x}^\varepsilon_\mathsf{s}(0), \mathbf{x}^\varepsilon_\mathsf{f}(0)) &= (\mathbf{x}^{\mathsf{init}}_\mathsf{s}, \mathbf{x}^{\mathsf{init}}_\mathsf{f}) \end{aligned} \tag{17}$$

and

$$\begin{aligned} d\mathbf{V}^\varepsilon_\mathsf{s}(t) &= [D_\mathsf{s}F_\mathsf{s}(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon_\mathsf{s}(t) + D_\mathsf{f}F_\mathsf{s}(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon_\mathsf{f}(t)]dt + \sum_{k\in\Lambda_\mathsf{s}} \mathsf{P}_\mathsf{s}\nu_k\sqrt{\beta_k(\mathbf{x}^\varepsilon(t))}dW_k(t) \\ d\mathbf{V}^\varepsilon_\mathsf{f}(t) &= \frac{1}{\varepsilon}[D_\mathsf{s}F_\mathsf{f}(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon_\mathsf{s}(t) + D_\mathsf{f}F_\mathsf{f}(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon_\mathsf{f}(t)]dt + \frac{1}{\sqrt{\varepsilon}} \sum_{k\in\Lambda_\mathsf{f}} \mathsf{P}_\mathsf{f}\nu_k\sqrt{\beta_k(\mathbf{x}^\varepsilon(t))}dW_k(t) \\ (\mathbf{V}^\varepsilon_\mathsf{s}(0), \mathbf{V}^\varepsilon_\mathsf{f}(0)) &= (\mathbf{v}^{\mathsf{init}}_\mathsf{s}, \mathbf{v}^{\mathsf{init}}_\mathsf{f}). \end{aligned} \tag{18}$$

In terms of components, we write $F_\mathsf{s} = (F_1, F_2, \dots, F_\ell)$ and $F_\mathsf{f} = (F_{\ell+1}, F_{\ell+2}, \dots, F_N)$.

Our goal is to derive, using the methods of singular perturbation theory, a reduced model for the equations (17), (18) as $\varepsilon \to 0$. We make the following additional assumptions regarding our basic system (17), (18):

[1] While the distinction between slow and fast reactions seems natural (the latter have large intensities), the same cannot be said for species. Our distinction here between slow and fast species is really a definition; we assume a dichotomy and then *define* a slow species as one whose population can only change through the occurrence of slow reactions, similarly for fast species.

Assumption 3.2. There is a smooth function $\mathscr{Z} : (\mathbb{R}^+)^\ell \to (\mathbb{R}^+)^{N-\ell}$ such that for each $\mathbf{x}_\mathsf{s} \in (\mathbb{R}^+)^\ell$, $F_\mathsf{f}(\mathbf{x}_\mathsf{s}, \mathscr{Z}(\mathbf{x}_\mathsf{s})) = \mathbf{0}$. Moreover, $\mathscr{Z}(\mathbf{x}_\mathsf{s})$ is the *unique*, globally attracting equilibrium solution for the ODE

$$\begin{aligned} \dot{\mathfrak{z}}_t(\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f}) &= F_\mathsf{f}(\mathbf{x}_\mathsf{s}, \mathfrak{z}_t(\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f})) \\ \mathfrak{z}_0(\mathbf{x}_\mathsf{s}, \mathbf{x}_\mathsf{f}) &= \mathbf{x}_\mathsf{f}. \end{aligned} \tag{19}$$

In other words, with the slow variable in (17) fixed at $\mathbf{x}_\mathsf{s}$, the fast variable converges to $\mathscr{Z}(\mathbf{x}_\mathsf{s})$ regardless of the initial condition.

Assumption 3.3. The matrix $D_\mathsf{f} F_\mathsf{f}(\mathbf{x})$ is uniformly negative definite.

Assumption 3.4. Let $\mathscr{S}(\mathbf{x})$ be the $(N-\ell) \times (N-\ell)$-matrix valued function on $\mathbb{R}^N$ with (i,j)-entry

$$\mathscr{S}_{ij}(\mathbf{x}) \overset{\text{def}}{=} \sum_{k=M_\mathsf{s}+1}^{M} \nu_k^{\ell+i} \nu_k^{\ell+j} \beta_k(\mathbf{x}) \qquad \text{for } 1 \le i,j \le N-\ell. \tag{20}$$

In words, if we use σ_f to denote the diffusion coefficient in the second line of (18) (i.e. the equation for $\mathbf{V}_\mathsf{f}^\varepsilon(t)$), then $\mathscr{S}$ corresponds to $\sigma_\mathsf{f}\sigma_\mathsf{f}^T$. We assume that

$$D_\mathsf{f} F_\mathsf{f}(\mathbf{x}) \mathscr{S}(\mathbf{x}) = \mathscr{S}(\mathbf{x}) D_\mathsf{f} F_\mathsf{f}(\mathbf{x})^T. \tag{21}$$

We need a little more notation and one more assumption to state our main result. Define $\mathbf{x}^0(t) \overset{\text{def}}{=} (\mathbf{x}_\mathsf{s}^0(t), \mathbf{x}_\mathsf{f}^0(t))$ where $\mathbf{x}_\mathsf{s}^0(t)$ solves the ODE

$$\begin{aligned} \dot{\mathbf{x}}_\mathsf{s}^0(t) &= F_\mathsf{s}(\mathbf{x}_\mathsf{s}^0(t), \mathscr{Z}(\mathbf{x}_\mathsf{s}^0(t))) \\ \mathbf{x}_\mathsf{s}^0(0) &= \mathbf{x}_\mathsf{s}^{\mathsf{init}} \end{aligned} \tag{22}$$

and

$$\mathbf{x}_\mathsf{f}^0(t) \overset{\text{def}}{=} \mathscr{Z}(\mathbf{x}_\mathsf{s}^0(t)). \tag{23}$$

Our last assumption is a form of *ergodicity* (see [PS08]) for the fast dynamics. For $\mathbf{v} = (v_1, v_2, \dots, v_N) \in \mathbb{R}^N$, $f = f(\mathbf{v}) \in C^1(\mathbb{R}^N)$, define

$$\begin{aligned} \nabla_\mathsf{s} f(\mathbf{v}) &= \left(\frac{\partial f}{\partial v_1}, \frac{\partial f}{\partial v_2}, \dots, \frac{\partial f}{\partial v_\ell} \right) \\ \nabla_\mathsf{f} f(\mathbf{v}) &= \left(\frac{\partial f}{\partial v_{\ell+1}}, \frac{\partial f}{\partial v_{\ell+2}}, \dots, \frac{\partial f}{\partial v_N} \right). \end{aligned} \tag{24}$$

For $f \in C^2(\mathbb{R}^N)$, let

$$(L_0 f)(t, \mathbf{v}) = \left\langle D_\mathsf{s} F_\mathsf{f}(\mathbf{x}^0(t)) \mathbf{v}_\mathsf{s} + D_\mathsf{f} F_\mathsf{f}(\mathbf{x}^0(t)) \mathbf{v}_\mathsf{f}, \nabla_\mathsf{f} f \right\rangle_{\mathbb{R}^{N-\ell}} + \frac{1}{2} \sum_{i,j=\ell+1}^{N} \alpha_{ij}^0(t) \frac{\partial^2 f}{\partial v_i \partial v_j} \tag{25}$$

where, for $\ell + 1 \le i, j \le N$,

$$\alpha_{ij}^0(t) \overset{\text{def}}{=} \sum_{k=M_\mathsf{s}+1}^{M} \nu_k^i \nu_k^j \beta_k(\mathbf{x}^0(t)). \tag{26}$$

The significance of the operator L_0 will be seen in subsection 4.2.

Assumption 3.5. For each $(t, \mathbf{v}_{\mathsf{s}}) \in \mathbb{R}^+ \times \mathbb{R}^\ell$ fixed, assume that the operator L_0 has one-dimensional null space characterized by

$$\begin{aligned} L_0 1(\mathbf{v}_{\mathsf{f}}) &= 0 \\ L_0^* \rho(t, \mathbf{v}_{\mathsf{s}}, \mathbf{v}_{\mathsf{f}}) &= 0 \end{aligned} \tag{27}$$

where $1(\mathbf{v}_{\mathsf{f}})$ denotes constants in $\mathbf{v}_{\mathsf{f}}$ and ρ is a certain invariant probability density whose form will be specified later. Here, L_0 and L_0^* are differential operators in $\mathbf{v}_{\mathsf{f}}$ with $t, \mathbf{v}_{\mathsf{s}}$ appearing as parameters.

For notational convenience, define the effective drift coefficient

$$B(t) \stackrel{\text{def}}{=} D_{\mathsf{s}} F_{\mathsf{s}}(\mathbf{x}^0(t)) - D_{\mathsf{f}} F_{\mathsf{s}}(\mathbf{x}^0(t))[D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t))]^{-1} D_{\mathsf{s}} F_{\mathsf{f}}(\mathbf{x}^0(t)). \tag{28}$$

Result 3.6 (Main Result)**.** As $\varepsilon \to 0$, the pair $(\mathbf{x}_{\mathsf{s}}^\varepsilon(t), \mathbf{V}_{\mathsf{s}}^\varepsilon(t))$ can be approximated by $(\mathbf{x}_{\mathsf{s}}^0(t), \mathbf{V}_{\mathsf{s}}^0(t))$ where $\mathbf{x}_{\mathsf{s}}^0(t)$ solves the limiting ODE (22) and $\mathbf{V}_{\mathsf{s}}^0(t)$ solves the limiting SDE

$$\begin{aligned} d\mathbf{V}_{\mathsf{s}}^0(t) &= B(t)\mathbf{V}_{\mathsf{s}}^0(t)dt + \sum_{k \in \Lambda_{\mathsf{s}}} \mathsf{P}_{\mathsf{s}} \nu_k \sqrt{\beta_k(\mathbf{x}^0(t))} dW_k(t) \\ \mathbf{V}_{\mathsf{s}}^0(0) &= \mathbf{v}_{\mathsf{s}}^{\mathsf{init}}. \end{aligned} \tag{29}$$

This will be established in the next section.

4. Singular Perturbations Analysis

In this section, we prove Result 3.6. First, in subsection 4.1, we establish that $\mathbf{x}_{\mathsf{s}}^\varepsilon(t)$ can be approximated by $\mathbf{x}_{\mathsf{s}}^0(t)$ as $\varepsilon \to 0$. Although this is a classical result in the theory of singular perturbations, we still describe the asymptotic expansions since they play a role in the ensuing stochastic analysis. We do, however, make the simplifying assumption that $\mathbf{x}_{\mathsf{f}}^{\mathsf{init}} = \mathscr{X}(\mathbf{x}_{\mathsf{s}}^{\mathsf{init}})$, i.e. the ODE starts on the slow manifold. This can be relaxed by including an initial layer in the perturbation calculations. Since our primary goal is to understand the asymptotic approximation of $\mathbf{V}_{\mathsf{s}}^\varepsilon(t)$ by $\mathbf{V}_{\mathsf{s}}^0(t)$, we freely use the aforementioned assumption to simplify the analysis.

Next, in subsections 4.2 and 4.3, we justify the approximation of $\mathbf{V}_{\mathsf{s}}^\varepsilon(t)$ by $\mathbf{V}_{\mathsf{s}}^0(t)$ using techniques similar to the *stochastic mode reduction* of [MTV], which in turn relies on a theorem of Kurtz ([Kur73], see also [Pap76], [PS08]). The method here is to approximate the solution of the backward equation for (18) by an *effective* backward equation, valid as $\varepsilon \to 0$. These calculations are described in subsection 4.2, where we compute the effective backward equation (56). Some supporting calculations involving the invariant probability density ρ (see Assumption 3.5) are spelled out in subsection 4.3.

We will repeatedly use Taylor's Theorem in the sequel: if $f : \mathbb{R}^N \to \mathbb{R}$ is smooth, then for $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$, $\mathbf{a} = (a_1, a_2, \ldots, a_N) \in \mathbb{R}^N$, we have

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^N \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})(x_i - a_i)(x_j - a_j) + R_2(\mathbf{x}, \mathbf{a})$$

where $R_2(\mathbf{x}, \mathbf{a})/\|\mathbf{x} - \mathbf{a}\|^2 \to 0$ as $\|\mathbf{x} - \mathbf{a}\| \to 0$.

4.1. **ODE reduction.** Write

$$\mathbf{x}^\varepsilon(t) = \mathbf{x}^0(t) + \varepsilon \mathbf{x}^1(t) + \varepsilon^2 \mathbf{x}^2(t) + \ldots$$

or, in terms of components,

$$x_i^\varepsilon(t) = x_i^0(t) + \varepsilon x_i^1(t) + \varepsilon^2 x_i^2(t) + \ldots \qquad \text{for } 1 \le i \le N.$$

For $1 \leq i \leq N$,

$$\begin{aligned} F_i(\mathbf{x}^\varepsilon(t)) &= F_i(\mathbf{x}^0(t)) + \sum_{k=1}^N \frac{\partial F_i}{\partial x_k}(\mathbf{x}^0(t))(x_k^\varepsilon(t) - x_k^0(t)) + \mathcal{O}(\varepsilon^2) \\ &= F_i(\mathbf{x}^0(t)) + \varepsilon \sum_{k=1}^N \frac{\partial F_i}{\partial x_k}(\mathbf{x}^0(t)) x_k^1(t) + \mathcal{O}(\varepsilon^2) \end{aligned}$$

Writing $\mathbf{x}_\mathsf{s}^\varepsilon(t) = \mathbf{x}_\mathsf{s}^0(t) + \varepsilon \mathbf{x}_\mathsf{s}^1(t) + \ldots$, $\mathbf{x}_\mathsf{f}^\varepsilon(t) = \mathbf{x}_\mathsf{f}^0(t) + \varepsilon \mathbf{x}_\mathsf{f}^1(t) + \ldots$, the equation (17) becomes

$$\begin{aligned} \dot{\mathbf{x}}_\mathsf{s}^0(t) + \mathcal{O}(\varepsilon) &= F_\mathsf{s}(\mathbf{x}_\mathsf{s}^0(t), \mathbf{x}_\mathsf{f}^0(t)) + \mathcal{O}(\varepsilon) \\ \dot{\mathbf{x}}_\mathsf{f}^0(t) + \mathcal{O}(\varepsilon) &= \frac{1}{\varepsilon} F_\mathsf{f}(\mathbf{x}_\mathsf{s}^0(t), \mathbf{x}_\mathsf{f}^0(t)) + \sum_{k=1}^N \frac{\partial F_\mathsf{f}}{\partial x_k}(\mathbf{x}_\mathsf{s}^0(t), \mathbf{x}_\mathsf{f}^0(t)) x_k^1(t) + \mathcal{O}(\varepsilon). \end{aligned} \tag{30}$$

Equating terms of order $1/\varepsilon$ in the second equation, we get

$$F_\mathsf{f}(\mathbf{x}_\mathsf{s}^0(t), \mathbf{x}_\mathsf{f}^0(t)) = 0,$$

and hence, by Assumption 3.2,

$$\mathbf{x}_\mathsf{f}^0(t) = \mathscr{X}(\mathbf{x}_\mathsf{s}^0(t)).$$

Equating terms of order 1 in the first equation, we see that $\mathbf{x}_\mathsf{s}^0(t)$ - the leading order approximation of $\mathbf{x}_\mathsf{s}^\varepsilon(t)$ - solves (22), as claimed.

4.2. **SDE reduction.** To facilitate the analysis, we start by fixing some notation. Let $\sigma^\varepsilon(t)$ be the $N \times M$ matrix given by

$$\sigma^\varepsilon(t) = \left[\nu_1 \sqrt{\beta_1^\varepsilon(\mathbf{x}^\varepsilon(t))} \quad | \quad \nu_2 \sqrt{\beta_2^\varepsilon(\mathbf{x}^\varepsilon(t))} \quad | \quad \ldots \quad | \quad \nu_M \sqrt{\beta_M^\varepsilon(\mathbf{x}^\varepsilon(t))} \right]. \tag{31}$$

Then, the SDE (12) can be written

$$\begin{aligned} d\mathbf{V}^\varepsilon(t) &= DF^\varepsilon(\mathbf{x}^\varepsilon(t))\mathbf{V}^\varepsilon(t)dt + \sigma^\varepsilon(t)dW(t) \\ \mathbf{V}^\varepsilon(0) &= \mathbf{v}^{\mathsf{init}}. \end{aligned} \tag{32}$$

Define $a^\varepsilon(t) \stackrel{\text{def}}{=} \sigma^\varepsilon(t)\sigma^\varepsilon(t)^T$. Then, for $1 \leq i, j \leq N$, the (i,j)-entry of $a^\varepsilon(t)$ is given by

$$a_{ij}^\varepsilon(t) = \sum_{k=1}^M \nu_k^i \nu_k^j \beta_k^\varepsilon(\mathbf{x}^\varepsilon(t)) \tag{33}$$

where $\nu_k = (\nu_k^1, \nu_k^2, \ldots, \nu_k^N)$. Note that ν_k^i represents the change in population of species S_i due to a single occurrence of reaction R_k. By Assumption 3.1, we have that $\nu_k^i = 0$ for $1 \leq i \leq \ell$ whenever $k \in \Lambda_\mathsf{f}$ and $\nu_k^i = 0$ for $\ell + 1 \leq i \leq N$ whenever $k \in \Lambda_\mathsf{s}$. Hence,

$$a_{ij}^\varepsilon(t) = \begin{cases} \sum_{k=1}^{M_\mathsf{s}} \nu_k^i \nu_k^j \beta_k(\mathbf{x}^\varepsilon(t)) & \text{for } 1 \leq i, j \leq \ell \\ \frac{1}{\varepsilon} \sum_{k=M_\mathsf{s}+1}^M \nu_k^i \nu_k^j \beta_k(\mathbf{x}^\varepsilon(t)) & \text{for } \ell + 1 \leq i, j \leq N \\ 0 & \text{else.} \end{cases} \tag{34}$$

Let $\alpha^\varepsilon(t)$ be the $N \times N$ matrix whose (i,j)-entry is given by

$$\alpha_{ij}^\varepsilon(t) \stackrel{\text{def}}{=} \begin{cases} a_{ij}^\varepsilon(t) & \text{if } 1 \leq i, j \leq \ell \\ \varepsilon a_{ij}^\varepsilon(t) & \text{if } \ell + 1 \leq i, j \leq N \\ 0 & \text{else.} \end{cases} \tag{35}$$

We are now in a position to define the backward Kolmogorov operators for the SDE (18). For $f = f(\mathbf{v}) \in C^2(\mathbb{R}^N)$, define the operators L^ε, A^ε, $\mathscr{L}^\varepsilon$ as follows:

$$
\begin{aligned}
(L^\varepsilon f)(t,\mathbf{v}) &\stackrel{\text{def}}{=} \langle D_{\mathsf{s}}F_{\mathsf{f}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{s}} + D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{f}}, \nabla_{\mathsf{f}} f\rangle_{\mathbb{R}^{N-\ell}} + \frac{1}{2}\sum_{i,j=\ell+1}^{N} \alpha_{ij}^\varepsilon(t)\frac{\partial^2 f}{\partial v_i \partial v_j} \\
(A^\varepsilon f)(t,\mathbf{v}) &\stackrel{\text{def}}{=} \langle D_{\mathsf{s}}F_{\mathsf{s}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{s}} + D_{\mathsf{f}}F_{\mathsf{s}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{f}}, \nabla_{\mathsf{s}} f\rangle_{\mathbb{R}^{\ell}} + \frac{1}{2}\sum_{i,j=1}^{\ell} \alpha_{ij}^\varepsilon(t)\frac{\partial^2 f}{\partial v_i \partial v_j} \\
(\mathscr{L}^\varepsilon f)(t,\mathbf{v}) &\stackrel{\text{def}}{=} \frac{1}{\varepsilon}(L^\varepsilon f)(t,\mathbf{v}) + (A^\varepsilon f)(t,\mathbf{v})
\end{aligned}
\tag{36}
$$

The backward equation for the SDE (18) is given by

$$
\frac{\partial w^\varepsilon}{\partial t}(t,\mathbf{v}) = (\mathscr{L}^\varepsilon w^\varepsilon)(t,\mathbf{v}). \tag{37}
$$

Writing

$$
w^\varepsilon = w_0 + \varepsilon w_1 + \varepsilon^2 w_2 + \dots, \tag{38}
$$

our goal is to identify and characterize the leading order term w_0 in terms of an *effective* backward equation. The SDE (29) has been chosen to yield precisely this effective backward equation, thus substantiating our claim in Result 3.6.

To start, note that the coefficients in the differential operators L^ε and A^ε, being functions of $\mathbf{x}^\varepsilon(t)$ exhibit dependence on *both* time and ε. To identify the dominant behavior as $\varepsilon \to 0$, we expand the operators L^ε and A^ε in powers of ε:

$$
\begin{aligned}
L^\varepsilon &= L_0 + \varepsilon L_1 + \varepsilon^2 L_2 + \dots \\
A^\varepsilon &= A_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \dots
\end{aligned}
\tag{39}
$$

For $1 \le i,j \le N$, we have

$$
\begin{aligned}
\frac{\partial F_i}{\partial x_j}(\mathbf{x}^\varepsilon(t)) &= \frac{\partial F_i}{\partial x_j}(\mathbf{x}^0(t)) + \sum_{k=1}^{N} \frac{\partial^2 F_i}{\partial x_k \partial x_j}(\mathbf{x}^0(t))(x_k^\varepsilon(t) - x_k^0(t)) + \mathcal{O}(\varepsilon^2) \\
&= \frac{\partial F_i}{\partial x_j}(\mathbf{x}^0(t)) + \varepsilon \sum_{k=1}^{N} \frac{\partial^2 F_i}{\partial x_k \partial x_j}(\mathbf{x}^0(t)) x_k^1(t) + \mathcal{O}(\varepsilon^2).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
D_{\mathsf{s}}F_{\mathsf{s}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{s}} &= D_{\mathsf{s}}F_{\mathsf{s}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + \varepsilon\sum_{k=1}^{N} x_k^1(t)\frac{\partial}{\partial x_k}(D_{\mathsf{s}}F_{\mathsf{s}})(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + \mathcal{O}(\varepsilon^2) \\
D_{\mathsf{f}}F_{\mathsf{s}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{f}} &= D_{\mathsf{f}}F_{\mathsf{s}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}} + \varepsilon\sum_{k=1}^{N} x_k^1(t)\frac{\partial}{\partial x_k}(D_{\mathsf{f}}F_{\mathsf{s}})(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}} + \mathcal{O}(\varepsilon^2) \\
D_{\mathsf{s}}F_{\mathsf{f}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{s}} &= D_{\mathsf{s}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + \varepsilon\sum_{k=1}^{N} x_k^1(t)\frac{\partial}{\partial x_k}(D_{\mathsf{s}}F_{\mathsf{f}})(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + \mathcal{O}(\varepsilon^2) \\
D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^\varepsilon(t))\mathbf{v}_{\mathsf{f}} &= D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}} + \varepsilon\sum_{k=1}^{N} x_k^1(t)\frac{\partial}{\partial x_k}(D_{\mathsf{f}}F_{\mathsf{f}})(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}} + \mathcal{O}(\varepsilon^2)
\end{aligned}
\tag{40}
$$

where the partial derivatives of the matrices $D_{\mathsf{s}}F_{\mathsf{s}}$, $D_{\mathsf{f}}F_{\mathsf{s}}$, $D_{\mathsf{s}}F_{\mathsf{f}}$, $D_{\mathsf{f}}F_{\mathsf{f}}$ are evaluated component-wise. For $1 \leq k \leq M$, expand $\beta_k(\mathbf{x}^\varepsilon(t))$ as follows:

$$\beta_k(\mathbf{x}^\varepsilon(t)) = \beta_k(\mathbf{x}^0(t)) + \varepsilon \sum_{m=1}^{N} \frac{\partial \beta_k}{\partial x_m}(\mathbf{x}^0(t)) x_m^1(t) + \mathcal{O}(\varepsilon^2). \tag{41}$$

Then, for $1 \leq i, j \leq \ell$, we have

$$\alpha_{ij}^\varepsilon(t) = \alpha_{ij}^0(t) + \varepsilon \sum_{k=1}^{M_{\mathsf{s}}} \sum_{m=1}^{N} \nu_k^i \nu_k^j \frac{\partial \beta_k}{\partial x_m}(\mathbf{x}^0(t)) x_m^1(t) + \mathcal{O}(\varepsilon^2) \tag{42}$$

where

$$\alpha_{ij}^0(t) \stackrel{\text{def}}{=} \sum_{k=1}^{M_{\mathsf{s}}} \nu_k^i \nu_k^j \beta_k(\mathbf{x}^0(t)) \tag{43}$$

and, for $\ell + 1 \leq i, j \leq N$, we have

$$\alpha_{ij}^\varepsilon(t) = \alpha_{ij}^0(t) + \varepsilon \sum_{k=M_{\mathsf{s}}+1}^{M} \sum_{m=1}^{N} \nu_k^i \nu_k^j \frac{\partial \beta_k}{\partial x_m}(\mathbf{x}^0(t)) x_m^1(t) + \mathcal{O}(\varepsilon^2) \tag{44}$$

where

$$\alpha_{ij}^0(t) \stackrel{\text{def}}{=} \sum_{k=M_{\mathsf{s}}+1}^{M} \nu_k^i \nu_k^j \beta_k(\mathbf{x}^0(t)). \tag{45}$$

We now have, that for $f \in C^2(\mathbb{R}^N)$,

$$(L_0 f)(t, \mathbf{v}) = \left\langle D_{\mathsf{s}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}}, \nabla_{\mathsf{f}} f \right\rangle_{\mathbb{R}^{N-\ell}} + \frac{1}{2} \sum_{i,j=\ell+1}^{N} \alpha_{ij}^0(t) \frac{\partial^2 f}{\partial v_i \partial v_j}. \tag{46}$$

While the explicit form of L_1 can be computed fairly easily, it will suffice for our analysis to note that L_1 is a linear operator that involves first and second partial derivatives in $v_{\ell+1}, v_{\ell+2}, \ldots, v_N$.

Also,

$$(A_0 f)(t, \mathbf{v}) = \left\langle D_{\mathsf{s}}F_{\mathsf{s}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}} + D_{\mathsf{f}}F_{\mathsf{s}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{f}}, \nabla_{\mathsf{s}} f \right\rangle_{\mathbb{R}^{\ell}} + \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_{ij}^0(t) \frac{\partial^2 f}{\partial v_i \partial v_j}. \tag{47}$$

Plugging $w^\varepsilon = w_0 + \varepsilon w_1 + \varepsilon^2 w_2 + \ldots$ into the backward equation, we get

$$\begin{aligned} \frac{\partial w_0}{\partial t} + \varepsilon \frac{\partial w_1}{\partial t} + \cdots &= \left(\frac{1}{\varepsilon} L_0 + L_1 + \varepsilon L_2 + \ldots \right) (w_0 + \varepsilon w_1 + \ldots) \\ &\quad + \left(A_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \ldots \right) (w_0 + \varepsilon w_1 + \ldots) \end{aligned} \tag{48}$$

Equating terms with like powers of ε yields

$$\begin{aligned} \mathcal{O}(1/\varepsilon): &\qquad L_0 w_0 &&= 0 \\ \mathcal{O}(1): &\qquad \frac{\partial w_0}{\partial t} &&= L_0 w_1 + L_1 w_0 + A_0 w_0, \qquad \text{etc.} \end{aligned} \tag{49}$$

The $\mathcal{O}(1/\varepsilon)$ equation implies that $w_0 \in \mathscr{N}(L_0)$. By Assumption 3.5, it follows that w_0 is *independent* of the fast variable $\mathbf{v}_{\mathsf{f}}$, i.e. $w_0 = w_0(t, \mathbf{v}_{\mathsf{s}})$. Since L_1 is a differential operator that involves partial derivatives with respect to the components of $\mathbf{v}_{\mathsf{f}}$, we also have $L_1 w_0 = 0$. Now, the $\mathcal{O}(1)$ equation becomes

$$L_0 w_1 = \frac{\partial w_0}{\partial t} - A_0 w_0. \tag{50}$$

Solvability of this equation requires that $\frac{\partial w_0}{\partial t} - A_0 w_0 \in \mathscr{R}(L_0)$, or equivalently, that $\frac{\partial w_0}{\partial t} - A_0 w_0 \in \mathscr{N}(L_0^*)^\perp$, where L_0^* denotes the formal adjoint of L_0. Assuming that $\mathscr{N}(L_0^*)$ is one-dimensional and is spanned by the probability density $\rho(t, \mathbf{v}_{\mathsf{s}}, \mathbf{v}_{\mathsf{f}})$, the effective backward equation is given by

$$\int_{\mathbf{v}_{\mathsf{f}} \in \mathbb{R}^{N-\ell}} \left(\frac{\partial w_0}{\partial t}(t, \mathbf{v}_{\mathsf{s}}) - (A_0 w_0)(t, \mathbf{v}_{\mathsf{s}}, \mathbf{v}_{\mathsf{f}}) \right) \rho(t, \mathbf{v}_{\mathsf{s}}, \mathbf{v}_{\mathsf{f}}) d\mathbf{v}_{\mathsf{f}} = 0. \tag{51}$$

The mode reduction is now completed by identifying ρ and computing explicitly the effective backward equation according to (51). We specify the correct ρ below and compute the effective backward equation and postpone the verification of $L_0^*\rho = 0$ to the next subsection.

For $t \geq 0$, let $\mu(t) \stackrel{\text{def}}{=} -[D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t))]^{-1} D_{\mathsf{s}} F_{\mathsf{f}}(\mathbf{x}^0(t)) \mathbf{v}_{\mathsf{s}}$. Let $S(t)$ be the $(N-\ell) \times (N-\ell)$ matrix with (i,j)-entry

$$S_{ij}(t) \stackrel{\text{def}}{=} \sum_{k=M_{\mathsf{s}}+1}^{M} \nu_k^{\ell+i} \nu_k^{\ell+j} \beta_k(\mathbf{x}^0(t)) \qquad \text{for } t \geq 0,\ 1 \leq i,j \leq N-\ell. \tag{52}$$

Note that $S(t) = \mathscr{S}(\mathbf{x}^0(t))$. Let $C(t)$ be the $(N-\ell) \times (N-\ell)$ symmetric, positive-definite matrix given by

$$C(t) \stackrel{\text{def}}{=} \int_0^\infty e^{s D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t))} S(t) e^{s D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t))^T} ds. \tag{53}$$

The uniform negative-definiteness of $D_{\mathsf{f}} F_{\mathsf{f}}$ ensures that the integral above converges. For each $t \geq 0$, $C(t)$ solves the *matrix Lyapunov equation*

$$D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t)) C(t) + C(t) D_{\mathsf{f}} F_{\mathsf{f}}(\mathbf{x}^0(t))^T = -S(t). \tag{54}$$

With $\mu = (\mu_{\ell+1}, \mu_{\ell+2}, \ldots, \mu_N)$ and $\mathbf{v}_{\mathsf{f}} = (v_{\ell+1}, v_{\ell+2}, \ldots, v_N)$, let

$$\rho(t, \mathbf{v}_{\mathsf{s}}, \mathbf{v}_{\mathsf{f}}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{(2\pi)^{N-\ell} \det(C)}} \exp\left\{ -\frac{1}{2} (\mathbf{v}_{\mathsf{f}} - \mu)^T C^{-1} (\mathbf{v}_{\mathsf{f}} - \mu) \right\}, \tag{55}$$

i.e. ρ is the Gaussian density with mean vector $\mu(t)$ and covariance matrix $C(t)$. Note that ρ depends on $\mathbf{v}_{\mathsf{s}}$ through $\mu(t)$. We will show in subsection 4.3 that $L_0^*\rho = 0$.

We can now finally provide the explicit form of the effective backward equation. Since ρ is a Gaussian probability density with mean $\mu(t)$ and covariance matrix $C(t)$, equation (51) becomes

$$\frac{\partial w_0}{\partial t}(t, \mathbf{v}_{\mathsf{s}}) = \left\langle D_{\mathsf{s}} F_{\mathsf{s}}(\mathbf{x}^0(t)) \mathbf{v}_{\mathsf{s}} + D_{\mathsf{f}} F_{\mathsf{s}}(\mathbf{x}^0(t)) \mu(t), \nabla_{\mathsf{s}} w_0(t, \mathbf{v}_{\mathsf{s}}) \right\rangle_{\mathbb{R}^\ell} + \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_{ij}^0(t) \frac{\partial^2 w_0}{\partial v_i \partial v_j}(t, \mathbf{v}_{\mathsf{s}}). \tag{56}$$

Recalling the expressions for $\mu(t)$ and $\alpha_{ij}^0(t)$, it is easily seen that (56) is the backward equation corresponding to the SDE (29).

4.3. **Verifying $L_0^*\rho = 0$.** We start by noting that since C is symmetric, $B \overset{\text{def}}{=} C^{-1}$ is symmetric as well. To simplify notation when taking partial derivatives of ρ, define, for $1 \le i \le N-\ell$,

$$\begin{aligned}\xi(i) &\overset{\text{def}}{=} -\frac{1}{2}\sum_{p=1}^{N-\ell}(b_{ip}+b_{pi})(v_{\ell+p}-\mu_{\ell+p}) \\ &= -\sum_{p=1}^{N-\ell} b_{ip}(v_{\ell+p}-\mu_{\ell+p})\end{aligned} \tag{57}$$

where we have used the fact that $B \overset{\text{def}}{=} (b_{ij})_{i,j=1}^{N-\ell}$ is symmetric. Straightforward computations yield that for $1 \le i,j \le N-\ell$,

$$\begin{aligned}\frac{\partial \rho}{\partial v_{\ell+i}} &= \xi(i)\rho \\ \frac{\partial^2 \rho}{\partial v_{\ell+j}\partial v_{\ell+i}} &= -b_{ij}\rho + \xi(i)\xi(j)\rho\end{aligned} \tag{58}$$

Recall that

$$(L_0 f)(t,\mathbf{v}) = \sum_{i=1}^{N-\ell}\left(\sum_{m=1}^{N}\frac{\partial F_{\ell+i}}{\partial x_m}(\mathbf{x}^0(t))v_m\right)\frac{\partial f}{\partial v_{\ell+i}}(\mathbf{v}) + \frac{1}{2}\sum_{i,j=1}^{N-\ell}S_{ij}(t)\frac{\partial^2 f}{\partial v_{\ell+j}\partial v_{\ell+i}}(\mathbf{v}). \tag{59}$$

To simplify the ensuing manipulations, we will use the notation

$$\mathsf{b}_{\ell+i}(t,\mathbf{v}) \overset{\text{def}}{=} \sum_{m=1}^{N}\frac{\partial F_{\ell+i}}{\partial x_m}(\mathbf{x}^0(t))v_m \qquad \text{for } 1 \le i \le N-\ell. \tag{60}$$

Now

$$\begin{aligned}(L_0^*\rho)(t,\mathbf{v}) &= -\sum_{i=1}^{N-\ell}\frac{\partial}{\partial v_{\ell+i}}\left[\mathsf{b}_{\ell+i}(t,\mathbf{v})\rho\right] + \frac{1}{2}\sum_{i=1}^{N-\ell}\sum_{j=1}^{N-\ell}S_{ij}(t)\frac{\partial^2\rho}{\partial v_{\ell+i}\partial v_{\ell+j}} \\ &= -\sum_{i=1}^{N-\ell}\frac{\partial}{\partial v_{\ell+i}}\mathbf{J}_i(t,\mathbf{v})\end{aligned} \tag{61}$$

where

$$\mathbf{J}_i(t,\mathbf{v}) \overset{\text{def}}{=} \mathsf{b}_{\ell+i}(t,\mathbf{v})\rho - \frac{1}{2}\sum_{j=1}^{N-\ell}S_{ij}(t)\frac{\partial\rho}{\partial v_{\ell+j}} \qquad \text{for } 1 \le i \le N-\ell. \tag{62}$$

We have

$$\begin{aligned}\mathbf{J}_i(t,\mathbf{v}) &= \mathsf{b}_{\ell+i}(t,\mathbf{v})\rho - \frac{1}{2}\sum_{j=1}^{N-\ell}S_{ij}(t)\xi(j)\rho \\ &= \mathbf{M}_i(t,\mathbf{v})\rho\end{aligned} \tag{63}$$

where

$$\mathbf{M}_i(t,\mathbf{v}) \overset{\text{def}}{=} \mathsf{b}_{\ell+i}(t,\mathbf{v}) - \frac{1}{2}\sum_{j=1}^{N-\ell}S_{ij}(t)\xi(j) \tag{64}$$

Thus

$$(L_0^*\rho)(t,\mathbf{v}) = -\sum_{i=1}^{N-\ell} \frac{\partial}{\partial v_{\ell+i}}\left[\mathbf{M}_i(t,\mathbf{v})\rho\right] \tag{65}$$

Recalling that $D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mu = -D_{\mathsf{s}}F_{\mathsf{f}}(\mathbf{x}^0(t))\mathbf{v}_{\mathsf{s}}$, it is easily seen that

$$\begin{aligned}\mathsf{b}_{\ell+i}(t,\mathbf{v}) &= \sum_{m=1}^{N} \frac{\partial F_{\ell+i}}{\partial x_m}(\mathbf{x}^0(t))v_m \\ &= \sum_{m=1}^{N-\ell} \frac{\partial F_{\ell+i}}{\partial x_{\ell+m}}(\mathbf{x}^0(t))(v_{\ell+m}-\mu_{\ell+m})\end{aligned} \tag{66}$$

Now

$$\begin{aligned}\mathbf{M}_i(t,\mathbf{v}) &= \sum_{m=1}^{N-\ell} \frac{\partial F_{\ell+i}}{\partial x_{\ell+m}}(\mathbf{x}^0(t))(v_{\ell+m}-\mu_{\ell+m}) + \frac{1}{2}\sum_{j=1}^{N-\ell}\sum_{m=1}^{N-\ell} S_{ij}(t)b_{jm}(v_{\ell+m}-\mu_{\ell+m}) \\ &= \sum_{m=1}^{N-\ell}(v_{\ell+m}-\mu_{\ell+m})\left\{\frac{\partial F_{\ell+i}}{\partial x_{\ell+m}}(\mathbf{x}^0(t)) + \frac{1}{2}\sum_{j=1}^{N-\ell} S_{ij}(t)b_{jm}\right\} \\ &= \sum_{m=1}^{N-\ell}(v_{\ell+m}-\mu_{\ell+m})\left\{(D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t)))_{im} + \frac{1}{2}(SB)_{im}\right\} \\ &= \left(\left(D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t)) + \frac{1}{2}SB\right)(\mathbf{v}_{\mathsf{f}}-\mu)\right)_i\end{aligned} \tag{67}$$

It is easily checked that the conditions

$$D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t)) + \frac{1}{2}S(t)B(t) = 0 \tag{68}$$

and

$$D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))C(t) = C(t)D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T \tag{69}$$

are equivalent. Recalling that (by Assumption 3.4)

$$D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))S(t) = S(t)D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T \tag{70}$$

and that for a square matrix A, $Ae^{At} = e^{At}A$, we get

$$\begin{aligned}D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))C(t) &= D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))\left(\int_0^\infty e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))}S(t)e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T}ds\right) \\ &= \int_0^\infty e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))}D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))S(t)e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T}ds \\ &= \int_0^\infty e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))}S(t)D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T}ds \\ &= \left(\int_0^\infty e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))}S(t)e^{sD_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T}ds\right)D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T \\ &= C(t)D_{\mathsf{f}}F_{\mathsf{f}}(\mathbf{x}^0(t))^T.\end{aligned} \tag{71}$$

Hence, $\mathbf{M}_i \equiv 0$ for $1 \leq i \leq N-\ell$, and thus $L_0^*\rho = 0$ as required.

5. Application

We now return to the example introduced in Section 2 to illustrate our results.

5.1. **Analytic expressions.** Let $k_r(p)$ be the Hill function given by

$$k_r(p) = \frac{k_r^{max}}{1 + (p/K_D)^N} \qquad \text{for } p \in \mathbb{R}. \tag{72}$$

The continuous-time Markov chain $(\mathsf{P}_n^\varepsilon(t), \mathsf{R}_n^\varepsilon(t))$ (corresponding to $\mathbf{X}^{n,\varepsilon}(t)$ in equation (4)) on $\mathbb{Z}^+ \times \mathbb{Z}^+$ that describes the population at time t of the protein and mRNA respectively, evolves according to

(73)

$$\mathbb{P}\Big((\mathsf{P}_n^\varepsilon(t+h), \mathsf{R}_n^\varepsilon(t+h)) = (\mathsf{P}_n^\varepsilon(t)+1, \mathsf{R}_n^\varepsilon(t))\Big|(\mathsf{P}_n^\varepsilon(s), \mathsf{R}_n^\varepsilon(s)) : 0 \le s \le t\Big) = k_p \mathsf{R}_n^\varepsilon(t) h + o(h)$$

$$\mathbb{P}\Big((\mathsf{P}_n^\varepsilon(t+h), \mathsf{R}_n^\varepsilon(t+h)) = (\mathsf{P}_n^\varepsilon(t)-1, \mathsf{R}_n^\varepsilon(t))\Big|(\mathsf{P}_n^\varepsilon(s), \mathsf{R}_n^\varepsilon(s)) : 0 \le s \le t\Big) = \gamma_p \mathsf{P}_n^\varepsilon(t) h + o(h)$$

$$\mathbb{P}\Big((\mathsf{P}_n^\varepsilon(t+h), \mathsf{R}_n^\varepsilon(t+h)) = (\mathsf{P}_n^\varepsilon(t), \mathsf{R}_n^\varepsilon(t)+1)\Big|(\mathsf{P}_n^\varepsilon(s), \mathsf{R}_n^\varepsilon(s)) : 0 \le s \le t\Big) = \frac{n}{\varepsilon} k_r \left(\frac{\mathsf{P}_n^\varepsilon(t)}{n}\right) h + o(h)$$

$$\mathbb{P}\Big((\mathsf{P}_n^\varepsilon(t+h), \mathsf{R}_n^\varepsilon(t+h)) = (\mathsf{P}_n^\varepsilon(t), \mathsf{R}_n^\varepsilon(t)-1)\Big|(\mathsf{P}_n^\varepsilon(s), \mathsf{R}_n^\varepsilon(s)) : 0 \le s \le t\Big) = \frac{1}{\varepsilon} \gamma_r \mathsf{R}_n^\varepsilon(t) h + o(h)$$

as $h \to 0$. As before, n is the reaction volume.

The characterization in terms of random time changes is

$$\begin{aligned} \mathsf{P}_n^\varepsilon(t) &= \mathsf{P}_n^\varepsilon(0) + Y_1\left(\int_0^t k_p \mathsf{R}_n^\varepsilon(s) ds\right) - Y_2\left(\int_0^t \gamma_p \mathsf{P}_n^\varepsilon(s) ds\right) \\ \mathsf{R}_n^\varepsilon(t) &= \mathsf{R}_n^\varepsilon(0) + Y_3\left(\frac{n}{\varepsilon}\int_0^t k_r\left(\frac{\mathsf{P}_n^\varepsilon(s)}{n}\right) ds\right) - Y_4\left(\frac{1}{\varepsilon}\int_0^t \gamma_r \mathsf{R}_n^\varepsilon(s) ds\right) \end{aligned} \tag{74}$$

where Y_i, $i = 1, 2, 3, 4$ are independent unit-rate Poisson processes. The macroscopic ODE approximating $(\mathsf{P}_n^\varepsilon(t)/n, \mathsf{R}_n^\varepsilon(t)/n)$ is

$$\begin{aligned} \dot{\mathsf{p}}^\varepsilon(t) &= k_p \mathsf{r}^\varepsilon(t) - \gamma_p \mathsf{p}^\varepsilon(t) \\ \dot{\mathsf{r}}^\varepsilon(t) &= \frac{1}{\varepsilon}\left[k_r(\mathsf{p}^\varepsilon(t)) - \gamma_r \mathsf{r}^\varepsilon(t)\right] \\ (\mathsf{p}^\varepsilon(0), \mathsf{r}^\varepsilon(0)) &= (\mathsf{p}_{\mathsf{init}}, \mathsf{r}_{\mathsf{init}}) \in \mathbb{R}^+ \times \mathbb{R}^+. \end{aligned} \tag{75}$$

The SDE approximating the fluctuations $\left(\sqrt{n}\left(\frac{\mathsf{P}_n^\varepsilon(t)}{n} - \mathsf{p}^\varepsilon(t)\right), \sqrt{n}\left(\frac{\mathsf{R}_n^\varepsilon(t)}{n} - \mathsf{r}^\varepsilon(t)\right)\right)$ as $n \to \infty$ takes the form

(76)

$$\begin{aligned} d\mathsf{V}^\varepsilon(t) &= \left[k_p \mathsf{U}^\varepsilon(t) - \gamma_p \mathsf{V}^\varepsilon(t)\right] dt + \sqrt{k_p \mathsf{r}^\varepsilon(t)} dW_1(t) - \sqrt{\gamma_p \mathsf{p}^\varepsilon(t)} dW_2(t) \\ d\mathsf{U}^\varepsilon(t) &= \frac{1}{\varepsilon}\left[k_r'(\mathsf{p}^\varepsilon(t)) \mathsf{V}^\varepsilon(t) - \gamma_r \mathsf{U}^\varepsilon(t)\right] dt + \frac{1}{\sqrt{\varepsilon}}\left[\sqrt{k_r(\mathsf{p}^\varepsilon(t))} dW_3(t) - \sqrt{\gamma_r \mathsf{r}^\varepsilon(t)} dW_4(t)\right] \\ (\mathsf{V}^\varepsilon(0), \mathsf{U}^\varepsilon(0)) &= (\mathsf{v}^{\mathsf{init}}, \mathsf{u}^{\mathsf{init}}) \in \mathbb{R} \times \mathbb{R} \end{aligned}$$

where $W(t) \stackrel{\text{def}}{=} (W_1(t), W_2(t), W_3(t), W_4(t))$ is a four-dimensional Brownian motion.

In the present context, Result 3.6 can be stated as follows: As $\varepsilon \to 0$, the pair $(\mathsf{p}^\varepsilon(t), \mathsf{V}^\varepsilon(t))$ can be approximated by $(\mathsf{p}^0(t), \mathsf{V}^0(t))$ where

$$\begin{aligned}
\dot{\mathsf{p}}^0(t) &= k_p \frac{k_r(\mathsf{p}^0(t))}{\gamma_r} - \gamma_p \mathsf{p}^0(t) \\
d\mathsf{V}^0(t) &= \left[k_p \frac{k_r'(\mathsf{p}^0(t))}{\gamma_r}\mathsf{V}^0(t) - \gamma_p \mathsf{V}^0(t)\right] dt + \sqrt{k_p \frac{k_r(\mathsf{p}^0(t))}{\gamma_r}} dW_1(t) - \sqrt{\gamma_p \mathsf{p}^0(t)} dW_2(t) \\
(\mathsf{p}^0(0), \mathsf{V}^0(0)) &= (\mathsf{p}_{\mathsf{init}}, \mathsf{v}_{\mathsf{init}})
\end{aligned} \tag{77}$$

The protein and mRNA dynamics thus decouple in the limit as $\varepsilon \to 0$, giving an approximate model for the protein dynamics alone via a closed set of equations.

Writing

$$\begin{aligned}
\mathsf{r}^\varepsilon &= \mathsf{r}^0 + \varepsilon \mathsf{r}^1 + \varepsilon^2 \mathsf{r}^2 + \dots \\
\mathsf{p}^\varepsilon &= \mathsf{p}^0 + \varepsilon \mathsf{p}^1 + \varepsilon^2 \mathsf{p}^2 + \dots,
\end{aligned}$$

straightforward calculations mimicking those in Subsection 4.1 yield the ODE in (77). The operators in the backward equation for the SDE (76) are given by (compare with equation (36))

$$\begin{aligned}
L^\varepsilon &= [k_r'(\mathsf{p}^\varepsilon(t))v - \gamma_r u]\frac{\partial}{\partial u} + \frac{1}{2}[k_r(\mathsf{p}^\varepsilon(t)) + \gamma_r \mathsf{r}^\varepsilon(t)]\frac{\partial^2}{\partial u^2} \\
A^\varepsilon &= [k_p u - \gamma_p v]\frac{\partial}{\partial v} + \frac{1}{2}[k_p \mathsf{r}^\varepsilon(t) + \gamma_p \mathsf{p}^\varepsilon(t)]\frac{\partial^2}{\partial v^2} \\
\mathscr{L}^\varepsilon &= \frac{1}{\varepsilon}L^\varepsilon + A^\varepsilon.
\end{aligned} \tag{78}$$

As before, the backward equation takes the form

$$\frac{\partial w^\varepsilon}{\partial t}(t,u,v) = (\mathscr{L}^\varepsilon w^\varepsilon)(t,u,v). \tag{79}$$

Expanding, we get

$$\begin{aligned}
L^\varepsilon &= L_0 + \varepsilon L_1 + \dots \qquad \text{where} \\
L_0 &= [k_r'(\mathsf{p}_0(t))v - \gamma_r u]\frac{\partial}{\partial u} + \frac{1}{2}[2k_r(\mathsf{p}_0(t))]\frac{\partial^2}{\partial u^2} \\
L_1 &= k_r''(\mathsf{p}_0(t))\mathsf{p}_1(t) v \frac{\partial}{\partial u} + \frac{1}{2}[k_r'(\mathsf{p}_0(t))\mathsf{p}_1(t) + \gamma_r \mathsf{r}_1(t)]\frac{\partial^2}{\partial u^2}, \qquad \text{etc.}
\end{aligned} \tag{80}$$

and

$$\begin{aligned}
A^\varepsilon &= A_0 + \varepsilon A_1 + \dots \qquad \text{where} \\
A_0 &= [k_p u - \gamma_p v]\frac{\partial}{\partial v} + \frac{1}{2}[k_p \mathsf{r}_0(t) + \gamma_p \mathsf{p}_0(t)]\frac{\partial^2}{\partial v^2} \\
A_1 &= \frac{1}{2}[k_p \mathsf{r}_1(t) + \gamma_p \mathsf{p}_1(t)]\frac{\partial^2}{\partial v^2}, \qquad \text{etc.}
\end{aligned} \tag{81}$$

With

$$w^\varepsilon = w_0 + \varepsilon w_1 + \varepsilon^2 w_2 + \dots, \tag{82}$$

the arguments of subsection 4.3 give

$$\int_{\mathbb{R}} \left(\frac{\partial w_0}{\partial t} - A_0 w_0\right)\rho = 0 \tag{83}$$

where $\rho \in \mathscr{N}(L_0^*)$.

Define

$$\rho(t,u,v) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u-m)^2}{2\sigma^2}\right] \tag{84}$$

where

$$m \stackrel{\text{def}}{=} \frac{k_r'(\mathsf{p}_0(t))v}{\gamma_r} \qquad \text{and} \qquad \sigma^2 \stackrel{\text{def}}{=} \frac{k_r(\mathsf{p}_0(t))}{\gamma_r}. \tag{85}$$

Note that assumption 3.4 is trivially satisfied and that

$$L_0^*\rho = \frac{\partial}{\partial u}[\gamma_r(u-m)\rho] + \gamma_r\sigma^2\frac{\partial^2\rho}{\partial u^2} = 0, \tag{86}$$

i.e. $\rho \in \mathscr{N}(L_0^*)$. Hence,

$$\int_{u\in\mathbb{R}} \left(\frac{\partial w_0}{\partial t}(t,v) - (A_0 w_0)(t,u,v)\right)\rho(t,u,v)du = 0. \tag{87}$$

The resulting PDE obtained by averaging (over $u \in \mathbb{R}$) the coefficients of A_0 with respect to the density ρ is

$$\frac{\partial w_0}{\partial t}(t,v) = \left(k_p\frac{k_r'(\mathsf{p}_0(t))v}{\gamma_r} - \gamma_p v\right)\frac{\partial w_0}{\partial v}(t,v) + \frac{1}{2}\left(k_p\frac{k_r(\mathsf{p}_0(t))}{\gamma_r} + \gamma_p\mathsf{p}_0(t)\right)\frac{\partial^2 w_0}{\partial v^2}(t,v). \tag{88}$$

Equation (88) is simply the backward Kolmogorov equation for the SDE for $\mathsf{V}^0(t)$ in the second line of equation (77).

5.2. **Numerical results.** Our numerical results were obtained using the MATLAB stiff solver `ode15s` and are summarized in figures 3 and 4. For the choice of parameter values $k_r^{max} = 10$, $\gamma_r = 1$, $k_p = 10$, $\gamma_p = 1$, $N = 4$, $K_D = 60$, we provide comparisons of $\mathsf{p}^\varepsilon(t)$ and $\mathsf{p}^0(t)$ in Figure 3 and comparisons of $\mathbb{E}[(\mathsf{V}^\varepsilon(t) - \mathbb{E}[\mathsf{V}^\varepsilon(t)])^2]$ and $\mathbb{E}[(\mathsf{V}^0(t) - \mathbb{E}[\mathsf{V}^0(t)])^2]$ in Figure 4, for various choices of ε.

Equations (75) and (76) describe a slow-fast ODE driving a slow-fast SDE. Once one solves the ODE to find $(\mathsf{p}^\varepsilon(t), \mathsf{r}^\varepsilon(t))$, the SDE for $(\mathsf{V}^\varepsilon(t), \mathsf{U}^\varepsilon(t))$ becomes a linear SDE (linear in the solution process) which can be solved explicitly. One can also write simple (time-inhomogeneous) ODE that describe the time evolution of the first and second moments of $(\mathsf{V}^\varepsilon(t), \mathsf{U}^\varepsilon(t))$. Our numerical results were obtained by integrating various ODE using `ode15s`.

Let's briefly describe these ODE. We follow here the presentation in Section 5.6 in [KS91]. Let [2]

$$\mathsf{X}^\varepsilon(t) \stackrel{\text{def}}{=} (\mathsf{V}^\varepsilon(t), \mathsf{U}^\varepsilon(t)), \qquad \mathsf{x}_{\mathsf{init}} \stackrel{\text{def}}{=} (\mathsf{v}_{\mathsf{init}}, \mathsf{u}_{\mathsf{init}}), \qquad A_\varepsilon(t) \stackrel{\text{def}}{=} \begin{bmatrix} -\gamma_p & k_p \\ \frac{k_r'(\mathsf{p}^\varepsilon(t))}{\varepsilon} & -\frac{\gamma_r}{\varepsilon} \end{bmatrix} \tag{89}$$

and

$$\sigma_\varepsilon(t) \stackrel{\text{def}}{=} \begin{bmatrix} \sqrt{k_p \mathsf{r}^\varepsilon(t)} & -\sqrt{\gamma_p \mathsf{p}^\varepsilon(t)} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{k_r(\mathsf{p}^\varepsilon(t))}}{\sqrt{\varepsilon}} & -\frac{\sqrt{\gamma_r \mathsf{r}^\varepsilon(t)}}{\sqrt{\varepsilon}} \end{bmatrix}. \tag{90}$$

Now

$$\begin{aligned} d\mathsf{X}^\varepsilon(t) &= A_\varepsilon(t)\mathsf{X}^\varepsilon(t)dt + \sigma_\varepsilon(t)dW_t \\ \mathsf{X}^\varepsilon(0) &= \mathsf{x}_{\mathsf{init}}. \end{aligned} \tag{91}$$

[2]We will sometimes indicate the ε dependence using a subscript, reserving the superscript for transposes and (matrix) inverses.

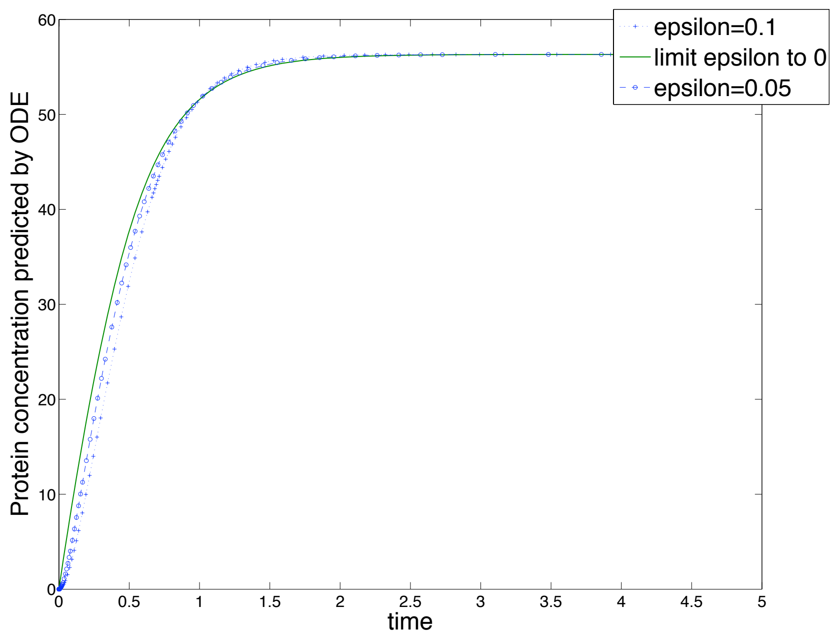

FIGURE 3. Comparison of $\mathsf{p}^0(t)$ (solid) and $\mathsf{p}^\varepsilon(t)$ for various choices of ε. The parameters in the problem were $k_r^{max} = 10$, $\gamma_r = 1$, $k_p = 10$, $\gamma_p = 1$, $N = 4$, $K_D = 60$.

Let $\Phi_\varepsilon(t)$ solve the *matrix* differential equation

$$\begin{aligned} \dot{\Phi}_\varepsilon(t) &= A_\varepsilon(t)\Phi_\varepsilon(t) \\ \Phi_\varepsilon(0) &= I \end{aligned} \tag{92}$$

where I denotes the 2×2 identity matrix. The process $\mathsf{X}^\varepsilon(t)$ is now given by

$$\mathsf{X}^\varepsilon(t) = \Phi_\varepsilon(t)\left[\mathsf{X}^\varepsilon(0) + \int_0^t \Phi_\varepsilon^{-1}(s)\sigma^\varepsilon(s)dW(s)\right], \qquad 0 \le t < \infty. \tag{93}$$

For the process $\mathsf{X}^\varepsilon(t)$, define the *mean vector* and *covariance matrix* functions

$$\begin{aligned} m_\varepsilon(t) &\stackrel{\text{def}}{=} \mathbb{E}[\mathsf{X}^\varepsilon(t)], \\ \rho_\varepsilon(s,t) &\stackrel{\text{def}}{=} \mathbb{E}[(\mathsf{X}^\varepsilon(s) - m_\varepsilon(s))(\mathsf{X}^\varepsilon(t) - m_\varepsilon(t))^T], \\ C_\varepsilon(t) &\stackrel{\text{def}}{=} \rho_\varepsilon(t,t) \end{aligned} \tag{94}$$

for $0 \le s, t < \infty$. Then, we have

$$\begin{aligned} \dot{m}_\varepsilon(t) &= A_\varepsilon(t)m_\varepsilon(t), \\ \dot{C}_\varepsilon(t) &= A_\varepsilon(t)C_\varepsilon(t) + C_\varepsilon(t)A_\varepsilon^T(t) + \sigma_\varepsilon(t)\sigma_\varepsilon^T(t). \end{aligned} \tag{95}$$

The approximate (limiting) model given by equation (77) also consists of an ODE (for $\mathsf{p}^0(t)$) driving an SDE (for $\mathsf{V}^0(t)$) - the SDE being linear (in $\mathsf{V}^0(t)$) and hence amenable to the type of

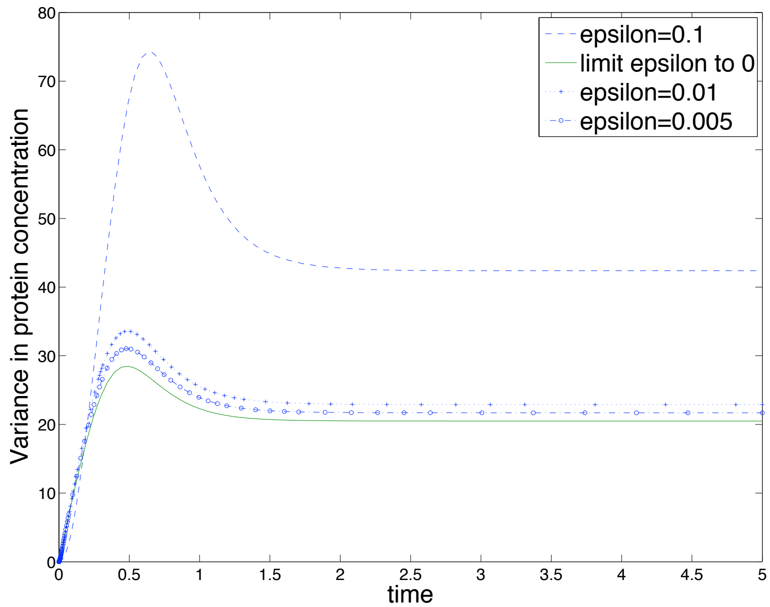

FIGURE 4. Comparison of $\mathtt{Var}(\mathsf{V}^0(t))$ (solid) and $\mathtt{Var}(\mathsf{V}^\varepsilon(t))$ for various choices of ε. The parameters in the problem were $k_r^{\max} = 10$, $\gamma_r = 1$, $k_p = 10$, $\gamma_p = 1$, $N = 4$, $K_D = 60$.

analysis described above. Indeed, define

$$\begin{aligned} \bar{A}(t) &\overset{\text{def}}{=} k_p \frac{k_r'(\mathsf{p}^0(t))}{\gamma_r} - \gamma_p, \\ \bar{\sigma}(t) &\overset{\text{def}}{=} \left[\sqrt{k_p \frac{k_r(\mathsf{p}^0(t))}{\gamma_r}} \quad - \sqrt{\gamma_p \mathsf{p}^0(t)} \right], \\ \bar{W}(t) &\overset{\text{def}}{=} \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}. \end{aligned} \tag{96}$$

Now,

$$\begin{aligned} d\mathsf{V}^0(t) &= \bar{A}(t)\mathsf{V}^0(t)dt + \bar{\sigma}(t)d\bar{W}(t) \\ \mathsf{V}^0(0) &= \mathsf{v}_{\mathsf{init}}. \end{aligned} \tag{97}$$

If we let

$$\bar{m}(t) \overset{\text{def}}{=} \mathbb{E}[\mathsf{V}^0(t)] \qquad \text{and} \qquad \bar{C}(t) \overset{\text{def}}{=} \mathbb{E}[(\mathsf{V}^0(t) - \mathbb{E}[\mathsf{V}^0(t)])^2] \tag{98}$$

then

$$\begin{aligned} \frac{d\bar{m}}{dt}(t) &= \bar{A}(t)\bar{m}(t) \\ \frac{d\bar{C}}{dt}(t) &= 2\bar{A}(t)\bar{C}(t) + (\bar{\sigma}\bar{\sigma}^T)(t) \end{aligned} \tag{99}$$

Note that $(\mathsf{V}^\varepsilon(t), \mathsf{U}^\varepsilon(t))$ and $\mathsf{V}^0(t)$ are time-inhomogeneous *Gaussian* processes (in $\mathbb{R}^2$ and $\mathbb{R}$ respectively) and hence are completely determined (in distribution) by their first two moments.

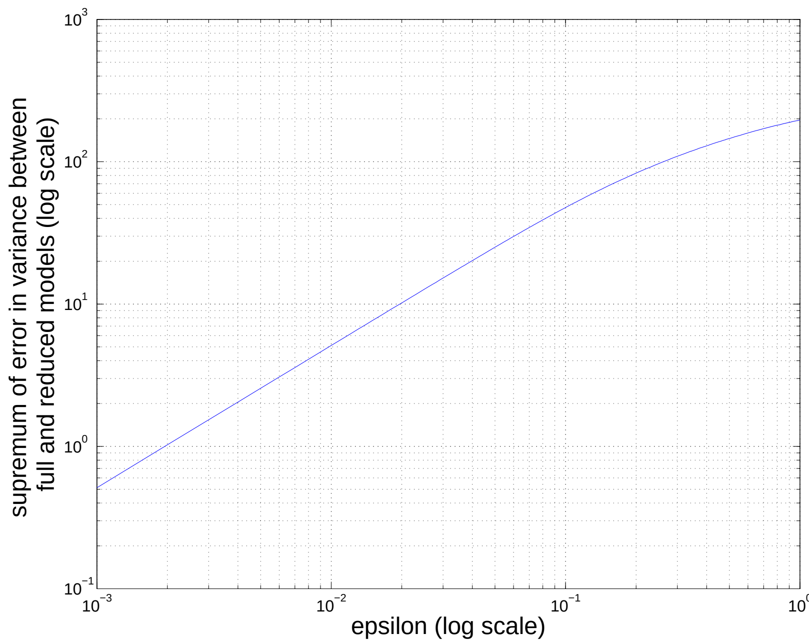

FIGURE 5. $\sup_{0\leq t\leq 5}|\texttt{Var}(\mathsf{V}^{\varepsilon}(t)) - \texttt{Var}(\mathsf{V}^{0}(t))|$ as a function of ε on a logarithmic scale. The parameters in the problem were $k_r^{\mathsf{max}} = 10$, $\gamma_r = 1$, $k_p = 10$, $\gamma_p = 1$, $N = 4$, $K_D = 60$.

References

[A07] Uri Alon. *An Introduction to Systems Biology. Design Principles of Biological Circuits*. Chapman & Hall/CRC.

[EE03] Johan Elf and Mans Ehrenberg. Fast Evaluation of Fluctuations in Biochemical Networks with the Linear Noise Approximation. *Genome Research*, 13:2475-2484, 2003.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons Inc., New York, 1986.

[Gar04] C. W. Gardiner. *Handbook of Stochastic Methods*, Springer, 2004.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Second Edition. Springer-Verlag, New York, 1991.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[Kur73] Thomas G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis*, 12:55-67, 1973.

[Kur78] Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6, no. 3, 223-240, 1978.

[MK06] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the chemical master equation. *Journal of Chemical Physics*, 124, 044104, 2006.

[MTV] Andrew J. Majda, Ilya Timofeyev, Eric Vanden Eijnden. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, 54, no.8, 891-974, 2001.

[Pap76] George C. Papanicolaou. Some probabilistic problems and methods in singular perturbations. *Rocky Mountain Journal of Mathematics*, Volume 6, Number 4, 653-674, 1976.

[PS08] Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Springer, 2008.

[RWA02] Christopher V. Rao, Denise M. Wolf and Adam P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, vol. 420, 231-237, 2002.

[SKPG] H. E. Samad, M. Khammash, L. Petzold and D. T. Gillespie. Stochastic modeling of gene regulatory networks. *Internat. J. Robust Nonlinear Control*, 15, 691-711, 2005.

[SCS03] Michael L. Simpson, Chris D. Cox and Gary S. Sayler. Frequency domain analysis of noise in autoregulated gene circuits. *Proceedings of the National Academy of Sciences*, vol. 100, no. 8, 4551-4556, 2003.

[vK] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*, 2001.